**Capstone 1 Project Proposal: Texas Education**

**What is the problem you want to solve?**

I would like to study the relationship between the percentage of college students who graduate within four years and the features of the school district they attended high school in. The study will focus on school districts located in the major regions of Texas (Houston, San Antonio, Dallas, Austin, Fort Worth) and Texas four-year colleges.

**Who is your client and why do they care about this problem?**

Parents who would like for their child to earn a college degree within four years. With the information from this study, parents could consider a move to or within the major regions. Earning a college degree will most likely open doors for their child and allow them to live a more comfortable life (what any concerned parent wants for their child). Graduating in four years also saves the family money when compared to a student who drops out or graduates in six years.

**What data are you using? How will you acquire the data?**

I will download district-level SAT, ACT, AP, and Wealth/ADA datasets from the Public Education Information Management System (“PEIMS”) on the Texas Education Agency’s website. Datasets on college enrollment and college graduation will be downloaded from the Texas Public Education Information Resource (“TPEIR”) website.

**How will you solve this problem?**

I will clean/wrangle all the datasets and then merge them together into one dataset that contains each school district’s historical features (SAT, ACT, AP, Wealth/ADA, College Enrollment) and the resulting percentage of students who earned their degree within four years. With historical feature and target data in place, I will attempt to build a model that can predict the percentage of students who will earn their degree within four years, based on the features of the school district they attended high school in.